DTU Library

Disentangling the Pure Time Effect From Site and Preference Heterogeneity Effects in Benefit Transfer: An Empirical Investigation of Transferability

Fetene, Gebeyehu Manie; Olsen, Søren Bøye; Bonnichsen, Ole

Published in:
Environmental and Resource Economics

Link to article, DOI:
10.1007/s10640-013-9751-y

Publication date:
2013

Document Version
Peer reviewed version

Disentangling the Pure Time Effect From Site and Preference Heterogeneity Effects in Benefit Transfer: An Empirical Investigation of Transferability

Gebeyehu M. Fetene, Søren B. Olsen & Ole Bonnichsen

Environmental and Resource Economics
The Official Journal of the European Association of Environmental and Resource Economists

ISSN 0924-6460

Environ Resource Econ
DOI 10.1007/s10640-013-9751-y

Springer

Environ Resource Econ
DOI 10.1007/s10640-013-9751-y

Disentangling the Pure Time Effect From Site and Preference Heterogeneity Effects in Benefit Transfer: An Empirical Investigation of Transferability

Gebeyehu M. Fetene · Søren B. Olsen · Ole Bonnichsen

Accepted: 22 November 2013

Abstract Two identical open-ended contingent valuation surveys assessing willingness-to-pay for better protection against flooding were administered in 2005 and 2010 at the same site. The 2010 survey was administered to the same respondents as those interviewed in 2005 as well as to new participants. This experimental design allows us to separate the pure temporal dimension from the spatial and social dimensions of transferability, thereby permitting an investigation of the temporal reliability associated with a transfer of value estimates over a 5-year time horizon. Having isolated the pure effect of time, the design further allows assessing the social dimension of transferability.

Keywords Open-ended contingent valuation · Benefit transfer · Transferability · Transfer error · Temporal reliability

1 Introduction

The primary motive for conducting benefit transfer (BT) is to avoid the substantial resource requirements associated with conducting a primary valuation study in terms of financial

S. B. Olsen (✉) · O. Bonnichsen
Department of Food and Resource Economics, Faculty of Science,
University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark
e-mail: sobo@ifro.ku.dk

O. Bonnichsen
e-mail: ole@ifro.ku.dk

G. M. Fetene
Department of Transport, Technical University of Denmark, Bygningstorvet 116B,
2800 Kgs. Lyngby, Denmark
e-mail: gebefe@transport.dtu.dk

Published online: 11 December 2013

costs, time[1] and decision delay costs.[2] The main drawback of using BT relates to the fact that transferring values from a study site to a policy site will inevitably introduce an error associated with the transfer. The sources of transfer error (TE) include any errors or mistakes made in the primary valuation survey at the study site from which values are transferred, differences between the study and policy sites not taken into account, choosing the wrong model or method of transfer, as well as mistakes made during the transfer process (Brouwer 2000; Colombo and Hanley 2008; Johnston and Rosenberger 2009; Baskaran et al. 2010). Despite this downside, it would seem reasonable to use BT as long as the survey cost savings are greater than the potential negative consequences of the TEs. Of course, this requires investigation of how reliable the transferred values are, and whether the cost savings of BT outweigh the consequences of the TEs.

A number of studies have been undertaken to test the reliability of BT and to investigate conditions under which BTs are applicable. So far, ambiguous results have been found (Johnston and Rosenberger 2009; Bateman et al. 2011). Some find that BT is unreliable (e.g. Rozan 2004), while others conclude that the reliability of BT depends on the similarity of the study and policy sites as well as the type of BT considered (Brouwer and Bateman 2005; Johnston and Rosenberger 2009; Baskaran et al. 2010; Bateman et al. 2011). Furthermore, the required level of precision can vary depending on the context that the BT is used. Previously BT has been used in many different contexts, including health (Barton 2002; Rozan 2004; Brouwer and Bateman 2005; Smith et al. 2006; Brouwer 2006), ecosystem services (Brouwer and Spaninks 1999; Baskaran et al. 2010), water quality (Muthke and Holm-Mueller 2004; Brouwer and Bateman 2005), private goods (Kealy et al. 1988), recreational areas (Loomis 1992; Downing and Ozuna 1996; Kirchhoff et al. 1997; Teisl et al. 1995) and amenities associated with house prices (Chattopadhyay 2003; Eshet et al. 2007).

There are also conflicting results about the relative performance of the different types of BT. Some studies find that benefit function transfer performs better than unit value transfers (Kirchhoff et al. 1997; Kaul et al. 2013), while others claim the opposite (Barton 2002; Källstrøm et al. 2010). Many results support that benefit function transfers perform relatively better for transfers between dissimilar sites, since it captures differences between the study and policy sites, while unit value transfers are as good as or even better suited for transfers between similar sites due to function transfers possibly over-parameterising similar sites (Chattopadhyay 2003; Brouwer and Bateman 2005; Bateman et al. 2011). However, there is no clear guideline to categorise sites as similar or dissimilar, and thus it is difficult to obtain a clear picture about the relative performance of BT types depending on site similarity.

In practice, BT typically involves transferring values in three dimensions: socially (across people), spatially (across sites) and temporally (across time). Any of these three dimensions could lead to TE and thus reduce the reliability and validity of BT. While the bulk of the literature has focused mainly on the social and spatial dimensions, there are relatively few BT studies that focus specifically on assessing the validity of transfers over time. The results from these few studies suggest that temporal BT is reliable for transfers over a short time intervals within 2 years (McConnell et al. 1998; Brouwer 2006; Bliem and Getzner 2012), but not for transfers over relatively longer time intervals of 5 years or more (Whitehead and Hoban 1999;

[1] While the costs of conducting primary evaluations are still substantial in terms of time spent, financial costs have decreased recently with the more widespread use of web-based surveys. However, conducting a BT can also be costly as it may require substantial effort from experts in the field.

[2] BT is sometimes the most viable option when quick valuation is required for urgent decision-seeking occasions. There could be insufficient time available to do primary valuation or data could be inaccessible (Johnston and Rosenberger 2009). For example, if a primary study is to be done to assess the impacts of an oil spill while oil is spilling, this would lead to substantial decision delay costs.

Brouwer and Bateman 2005; Zandersen et al. 2007). However, the BT studies focusing on temporal reliability, and especially those testing a transfer across more than 2 years (which will often be very relevant in practical BT for policy purposes) are few in number and several aspects have yet to be investigated thoroughly (Johnston and Rosenberger 2009; Bliem and Getzner 2012). Particularly, temporal BT studies that use the open-ended contingent valuation method (CVM) for transfers over a relatively long time interval (e.g. 5 years) where no major events likely to markedly affect willingness-to-pay (WTP) have occurred, have yet to be investigated (Johnston and Rosenberger 2009). Some of the previous temporal BT studies used revealed preference valuation methods (Zandersen et al. 2007), and those using CVM surveys were either not from open-ended CVM surveys or used relatively short time intervals of <2 years (e.g. Loomis 1989; Cameron 1997). Moreover, while some of the previous studies exclude the spatial dimension of the TE by using the same study and policy site, most of them have considered transfers across time while simultaneously transferring across individuals. Hence, the social and the temporal dimension of TE are confounded, and it is thus not possible to conjecture to what extent an observed transfer error is caused by a pure time effect or preference heterogeneity (Zandersen et al. 2007; León and Araña 2012), see for example studies by Whitehead and Hoban (1999), Brouwer (2006) and Brouwer and Bateman (2005).

In this paper, we contribute to the limited number of temporal BT studies in general, and in particular we assess temporal reliability of BT for a transfer across a long time interval of 5 years, using two open-ended CVM surveys concerning WTP for flood risk reductions conducted at the same site in the years 2005 and 2010. We transfer values both across identical respondents in a test–retest setting as well as across different respondents. This experimental design allows us to assess the social and temporal dimension of transferability separately, based on two different transfers. The first transfer is conducted purely in the temporal dimension[3] by transferring values from year 2005 to year 2010 for identical respondents, i.e. using a test–retest setting. This isolates the pure time effect and thus the actual temporal reliability, since both respondents and the site being valued in the two surveys are identical. To our knowledge, we are the first to assess the pure time effect over such a relatively long interval. The second transfer is conducted from the 2005 survey to a new sample of respondents recruited for the 2010 survey. Thus, the second transfer represents a transfer both in the temporal and the social dimension. Compared to the first transfer, any changes in TE should thus be attributable to the addition of the social dimension. While this split sample setup is similar to that of McConnell et al. (1998) where a short time interval of 2 months is tested using a willingness-to-accept (WTA) valuation question, we contribute to the BT literature by testing what is arguably more relevant for practical application of BT in general, namely a 5-year time interval and a WTP setting. Moreover, we assess the transferability differences between the three commonly used types of BT (unadjusted unit value, adjusted unit value and benefit function transfers) in terms of TE magnitudes as well as a range of statistical BT tests. Using statistical best-fit models, we conduct benefit function transfers both based on the full information available from the datasets as well as a more practically oriented approach based only on information that would have been available had the 2010 survey not been conducted. We also assess whether there are improvements in transferability performance when only theory-derived variables are used in the benefit function transfer, as suggested by Brouwer and Bateman (2005), Brouwer (2006) and Bateman et al. (2011). Finally, in the test–retest sample we investigate which factors determine intra-respondent change in WTP over the 5-year time span.

[3] While we refer to time as one dimension, we acknowledge that the temporal dimension is multifaceted in the sense that temporal changes may include both observable (e.g. age, income, household size) as well as latent (e.g. perceptions, attitudes, knowledge) factors.

The paper is organised as follows: Sect. 2 describes BT in more detail and reviews the literature on temporal BT. Section 3 describes the experimental setup of our survey as well as the methods and tests used to assess temporal reliability. Section 4 reports the results and provides some discussion, while Sect. 5 contains concluding remarks.

2 Benefit Transfer

2.1 Background

Benefit transfer (BT), also known as value transfer, is the transfer of valuation results from one or more previous primary studies of a site to another site or to the same site at a different point of time for which there is limited data (Brouwer 2000; Kristofersson and Navrud 2005; Navrud and Ready 2007). While the former is referred to as the study site, the latter is referred to as the policy site since some policy or project is considered there and value estimates are thus needed. Transferable values can be unit values, predicted values from functions or the functions themselves as well as a synthesis of many studies[4] (Bateman et al. 2000; Pearce et al. 2006; Navrud and Ready 2007).

2.1.1 Unit Value Transfer

The simple unit value transfer is the practice of directly transferring the average values of the study site to policy site without adjusting for differences between sites, implicitly assuming that valuation results are identical at the sites. Thus, the unadjusted unit value transfer is simply:

$$\overline{WTP_s} = \overline{WTP_p} \tag{1}$$

where $\overline{WTP_s}$ and $\overline{WTP_p}$ are the average WTPs of the study and policy sites respectively. While it is the simplest way to conduct BT in practice, its main limitation is the rather strong implicit assumption of similarity between the study and policy sites (Bateman et al. 2002).

Relaxing this assumption somewhat, the adjusted unit value transfer approach permits taking basic differences between the study and policy sites into account by multiplying $\overline{WTP_s}$ with relevant adjustment factors. Bateman et al. (2002) categorises adjustment factors into socio-economic and demographic differences of the relevant populations, differences in physical characteristics between sites, and differences in the proposed change and market conditions of the two sites. The adjusted unit value transfer is computed as:

$$\overline{WTP_p} = \overline{WTP_s}\left(\frac{A_p}{A_s}\left[\frac{Y_p}{Y_s}\right]^e\right) \tag{2}$$

where A denotes non-income adjustable differences, Y is income and e is the WTP elasticity of income at the study site for the resource in question. The justification behind the adjusted unit value transfer is that as long as basic differences between the two sites are controlled for, the transfer of average values should work. However, in practice unit values are typically adjusted only for income differences, known as income-adjusted unit value transfer. While easy to utilise, its main limitation is the implicit assumption that the income adjustment controls for all WTP differences between the study and policy sites while in fact differences in physical features of the valued resources, availability of substitutes, other socio-demographic

[4] Since we do not use meta-analysis function transfer in this paper, we do not provide a detailed account of this approach here.

conditions of the relevant population and market conditions may not be accounted for by the simple income adjustment (Bateman et al. 2000, 2002; Pearce et al. 2006; Navrud and Ready 2007; Baskaran et al. 2010).

2.1.2 Benefit Function Transfer

The second category of BT is function transfer, where the estimated coefficients are transferred from the study site to the policy site. The justification behind function transfer is that once the main covariates explaining WTP are controlled for, there will be less difference between primary and transferred estimates. The underlying assumption here is that the covariates explaining WTP and their explanatory power are identical at the policy and study sites—an assumption which is often regarded as the main limitation of function transfers (Bateman et al. 2000, 2002; Pearce et al. 2006; Navrud and Ready 2007; Baskaran et al. 2010). For example, suppose that WTP estimated from an open-ended CVM at a study site can be described as a simple linear function:

$$WTP_s = a_0 + a_1 G_s + a_2 H_s \tag{3}$$

where s indicates study site, a_0 to a_2 are coefficients to be estimated (a_1 and a_2 are row vectors), G represents site characteristics and H denotes household characteristics. A benefit function transfer from the above equation takes the estimated coefficients (a_0 to a_2) from the regression equation and substitutes the average values of explanatory variables (G and H) from the policy site census or other reliable data sources. The resulting transferred WTP can be described as:

$$WTP_p = a_0 + a_1 G_p + a_2 H_p \tag{4}$$

where p denotes the policy site; a_0 to a_2 are coefficients transferred from the study site; G_p and H_p are average values of site and household characteristics at the policy site (Barton 1999). Conducting function transfer requires availability of mean and other values of explanatory variables from existing datasets at the policy site. The values of some of the explanatory variables, e.g. socio-demographic variables, may be obtained from municipalities, census and other data collecting sources. Values of other variables like attitudinal or perception variables, however, may not be available from existing data sources.

2.2 Temporal Benefit Transfer Literature

It is common for decision makers to use previous study results as input for current decisions, and often over a considerable time span. For instance, Zandersen et al. (2007) reports that the Natural Resources Conservation Service in USA has used temporal BT more than 1,000 times within 15 years to valuate recreational impacts of small projects. These and other similar temporal BTs presume that preferences, market conditions and other determinants of the interest variable are fairly stable over time. The question whether these presumptions are correct is an empirical issue. Problems of obtaining additional and sufficient data years after the original study as well as methodological advances and developments in data collection techniques over time, often impede proper investigations of this issue. Hence, the literature investigating temporal transferability of valuation results is somewhat limited (Johnston and Rosenberger 2009; Bliem and Getzner 2012). McConnell et al. (1998) provides an overview of ten different temporal reliability CVM studies conducted prior to 1998. In Table 1, we supplement McConnell et al. (1998)'s overview with the temporal reliability CVM studies that have been published since 1998 as well as earlier studies that were not

Table 1 Summary of temporal reliability CVM studies not included in McConnell et al. (1998)

Author(s)	Question format	Time interval	Identical respondents	Good	Main findings concerning temporal reliability
Bliem and Getzner (2012)	Payment card	1 year	No	River restoration	Mean WTP values and preferences are stable While both projected and actual values are similar after removing outliers, equality of WTP values obtained from benefit functions is rejected
Brouwer (2006)	Dichotomous choice and open-ended	9 months	No	Reducing bathing risk associated with the 2003 bad weather in Europe	No significant difference in mean WTP Theory-derived functions are transferable, but the statistically best-fit function is not
Brouwer and Bateman (2005)	Dichotomous choice	5 years	No	Flood control of recreational and wetland area	Real WTP declines over time Mean equality of WTP rejected Theory-derived functions are transferable, but the statistically best-fit function is not
Onwujekwe et al. (2005)	Bidding game, binary with follow up and haggling technique	Within a month	Yes	Private good (insecticide-treated mosquito nets)	CVM is reliable and the three question formats have a similar level of reliability
Berrens et al. (2000)	Dichotomous choice	1 year	No	Protecting in-stream river flows	WTP functions are temporally stable
Whitehead and Hoban (1999)	Referendum	5 years	No	Reducing air and water pollution	WTP declines over time After controlling for attitude, WTP is stable Attitudes change over time

Table 1 continued

Author(s)	Question format	Time interval	Identical respondents	Good	Main findings concerning temporal reliability
McConnell et al. (1998)	Dichotomous choice	2 months	Yes	Angling	Mean and median WTA, preference and determinants of utility are stable over time
Cameron (1997)	Open-ended	1–2 years	Yes	River restoration (water quality)	Mean WTP did not change significantly over time The bid function remained stable over time
Downing and Ozuna (1996)	Referendum	From 1 to 3 years	No	Angling saltwater	Many of the benefit functions are transferable, but not the estimated welfare measure (median WTP) Function transfer overestimates WTP
Non-CVM temporal reliability studies					
Bliem et al. (2012)	Choice experiment	1 year	No	River restoration	Preferences and WTP estimates are not sensitive to time Individual preferences are robust over a short time period unless there are extreme events
Liebe et al. (2012)	Choice experiment	11 months	Yes	Onshore wind power externality	Overall the estimates are moderately stable over time
Bhattacharjee et al. (2009)	Travel cost method	2 years	Yes	Lake recreation	Overall the estimates are fairly stable over the study years
Parsons and Stefanova (2009)	Travel cost method	7 years	No	Beach trip	Preferences show qualitative stability over time Estimated coefficients and measures of WTP are significantly different over time
Skjoldborg et al. (2009)	Choice experiment	4 months	Yes	Private good (TNF-alpha inhibitors)	Observed a temporal reliability of the choice experiment
Zandersen et al. (2007)	Travel cost method	20 years	No	Recreational service of forests	Benefit functions are not transferable over longer years Preferences changed over time

included by McConnell et al. (1998). We furthermore add some recent non-CVM studies for comparison.

As can be seen in Table 1, the time elapsed between the studies at the study site and policy site used to assess transferability ranges from 2 weeks (Kealy et al. 1988) to 20 years (Zandersen et al. 2007). Table 1 suggests that temporal BT is stable over shorter time periods (Kealy et al. 1988; Loomis 1989; Carson et al. 1997; Brouwer 2006), but results concerning temporal reliability are more ambiguous when looking at longer time spans of more than 1 year (Whitehead and Hoban 1999; Brouwer and Bateman 2005; Zandersen et al. 2007). The majority of the studies are from dichotomous choice or referendum-type CVM surveys and there is a lack of studies from open-ended CVMs for transfers over relatively longer time periods. Table 1 also shows an observed decline of WTP over time for non-transferable findings (Brouwer and Bateman 2005; Zandersen et al. 2007). This may be somewhat unexpected since there is generally increasing scarcity of environmental resources over time (Navrud 2007). However, many other aspects could affect WTP, and it would seem inappropriate to generalise based on the current number of studies. Turning to each of the papers in more detail, it is evident that the topics addressed, issues treated and approaches taken in the literature are far from homogeneous and coordinated.

Recently, Bliem and Getzner (2012) assessed temporal stability of WTP values for river restoration in Austria using payment card CVMs conducted in 2007 and 2008. From this 1-year-interval BT study, they found temporal stability of mean WTP values as well as preferences after removing outliers. They also reported that estimated WTP bids were roughly in the same order of magnitude though statistical equality of WTP estimated from these bid functions was rejected. Bliem et al. (2012) also found similar results from the same river restoration valuation, but using the choice experiment method. Both the underlying indirect utility functions and marginal WTP for different attributes of the choice experiment were not significantly different over the study period. However, the reported results in both studies may not show the pure time effect since the transfer was not over identical respondents.

Another temporal BT study that is unable to disentangle the pure time effect from preference heterogeneity is Brouwer and Bateman (2005). They assessed temporal BT in a study of WTP for flood control and wetland conservation using dichotomous choice CVM surveys from 1991 and 1996. The main finding was that real WTP declined over time. Moreover, they found that the statistically best-fit benefit function was not transferable, but restricting the benefit function in terms of using only theory-derived factors made it transferable. Similar findings have also been observed from another 5-year-interval temporal BT study: Whitehead and Hoban (1999) who estimated WTP to reduce air and water pollution, found that mean WTP had declined over a 5-year period in a between-respondent study. However, the two mentioned studies differ in some aspects. While Whitehead and Hoban (1999) found that a change in attitude over time was responsible for an observed change of bid functions over time, thereby implying that including attitudinal variables improves temporal transferability, Brouwer and Bateman (2005) reported that including such attitudinal and other non-theory-derived variables impedes temporal transferability.

Brouwer (2006) used dichotomous choice CVM surveys and analysed the transferability of WTP values for improving bathing water quality. The author used two surveys conducted in 2002 and 2003 and assessed transferability of WTP values. The paper found similar results as Brouwer and Bateman (2005), in that a benefit function using only theory-derived factors is transferable. However, the results from this study cannot be attributed solely to a pure time effect since the respondents in the two surveys were not identical.

Onwujekwe et al. (2005) also investigated temporal reliability of CVM and compared three elicitation techniques, namely bidding game, binary (dichotomous) with a follow up question and finally a haggling technique. In their 1-month-interval assessment of temporal reliability of CVM for a marketed good, they concluded that CVM is reliable and that the three question formats have similar levels of reliability. The results from this study could be analysed on a pure time effect basis, since the same individuals were interviewed in the two periods. However, a carry-over effect was likely since the time elapsed between the two interviews was less than a month. From another short time interval study, McConnell et al. (1998) found a similar result for the temporal reliability of dichotomous CVM results, in that preferences for giving up sport fishing, mean WTA values and estimated benefit functions remain stable after 2 months in a pure time effect test–retest study.

Berrens et al. (2000) also assessed temporal reliability of dichotomous CVM when estimating benefits of river in-stream flows in Mexico. Using both Wald and likelihood ratio tests, they were unable to reject the null hypothesis of equality of the two estimated statistical best-fit benefit functions obtained from 1-year-interval surveys from 1995 and 1996. Unlike the results obtained by Brouwer (2006) and Brouwer and Bateman (2005), the statistical best-fit functions that included attitudinal and preference variables were found to be transferable. However, since the respondents in the two surveys were not identical, the results from their study cannot be attributed solely to a pure time effect.

The only open-ended format CVM temporal reliability studies are Cameron (1997) and Loomis (1989). Cameron (1997) valued water quality improvements in Australia using two groups of respondents; 'the research group' that consisted of government officials, farmers and other selected volunteers who had direct interest in the river, and 'the reference group' that were randomly selected from the population. The results suggest that mean WTP was stable over a 2-year time interval. However, a carry-over effect is highly likely since the reference group samples had been participating in monthly meetings about the study variable. Moreover, even though the study was able to investigate a pure time effect by resurveying identical respondents, the analyses rely on a small sample size of only 14 respondents validly participating in all CVM surveys. The other study using open-ended CVM is Loomis (1989) who investigated lake preservation using two surveys conducted with a 9-month interval. He also found stable WTP over time and statistically insignificant differences between the two periods' benefit functions.

Turning to the non-CVM studies shown in the lower part of Table 1, the results of these are similar to the CVM studies. Specifically, the valuation results are stable over short time periods but not over relatively longer time periods. Using a revealed preference approach, specifically the travel cost method, Zandersen et al. (2007) and Parsons and Stefanova (2009) assessed temporal transferability of WTP for car-borne forest recreation and beach trips, respectively. The two studies were conducted using time intervals of 20 and 7 years, but neither of the studies investigated pure time effects by using identical respondents. The studies found both mean values as well as estimated benefit functions to be non-transferable. On the other hand, in a choice experiment survey concerning negative externalities of onshore wind power, Liebe et al. (2012) found so-called moderately stable values from two 11-month-interval studies using identical respondents. Other non-CVM studies, which assessed temporal reliability of values over time with survey gaps ranging from 4 months to 2 years, also generally found time insensitive valuation results (Skjoldborg et al. 2009; Bhattacharjee et al. 2009; Bliem et al. 2012).

3 Methods

3.1 Survey and Questionnaire

This study is based on two identical open-ended CVM surveys conducted in 2005 and 2010 both concerning flood risk preferences. Both surveys were conducted in the same geographical area along the western coast of Jutland in Denmark. The targeted population was defined as people residing in areas that are <5 m above sea level. These areas are particularly prone to flooding from the North Sea during severe storms when the tide is high, and parts of these areas have previously experienced flooding.[5] A proper sampling frame was constructed by combining GIS-based topographical information with residential location information. A paper-and-pen-based self-administered open-ended CVM questionnaire was sent via ordinary mail to 1,413 randomly selected individuals from the sampling frame, aged 18 years or above in 2005. An identical questionnaire was administered in 2010 to all the respondents from 2005 as well as to 478 new respondents that had not participated in the 2005 survey.

The surveys included valuation questions concerning a proposed project that would result in a reduction of the risk of being flooded from the current risk of once in 100 years[6] to a lower risk level.[7] Specifically, respondents were asked to state how much they were willing to pay in terms of a yearly lump-sum payment to a private firm for installing additional flood protection measures that would lead to the described risk reductions.

The response rates were 77 % in 2005 and 54 % in 2010. After removing incomplete responses and protest zero bidders, the effective sample sizes relevant for this study were 581 in 2005 and 390 in 2010. Out of these respondents, 150 validly participated in both the 2005 and 2010 surveys. These 150 respondents are used as the main sample to assess the pure temporal dimension of BT transferability. The BT over time for these respondents can be used to assess the temporal transferability of BT since any differences in value estimates cannot be

[5] Historically, hundreds of lives have been lost due to flooding in the area. However, as a result of dike constructions and, more recently, a storm surge warning management system (and probably more clever residential location choices) it has been more than 100 years since flooding in the area has caused human fatalities. The most recent significant storm surges in the area were in 1976 and 1981, where large areas were flooded. All residents were evacuated from the areas in time, but damages to buildings and residences were severe and significant amounts of agricultural crops and production animals were lost.

[6] At the time of the survey, the flood protection measures taken by the Danish Coastal Authority in the area all aimed at the official target of limiting flooding events to cases where a so-called 100-year storm or worse occurred.

[7] To test for scope effects, respondents were randomly selected into two split-samples in the surveys. The first split-sample was presented with a project proposal that targeted the reduction in risk of flooding from the current once in 100 years to once in 200 years and the other sample was presented a project that targeted a risk reduction from the current once in 100 years to once in 500 years. In the 2005 survey, simple mean WTP was slightly higher for the 500-year scenario than the 200-year scenario which is in accordance with the theoretically predicted scope effect. However, the difference was not statistically significant. Furthermore, a similar comparison in the 2010 survey revealed a slightly lower WTP for the 500-year scenario than the 200-year scenario. This could indicate a scope issue but, again, the difference was not significant. In the background reports for these data (Tranberg et al. 2005; Dubgaard et al. 2011) it was argued that diminishing—and close to zero—marginal utility of additional risk reductions is the likely explanation for the lack of significant increase in WTP, since respondents find it hard to distinguish between the relatively low risk levels. Whether embedding bias or marginal utility close to zero is the explanation, for the present paper this issue is only of minor importance since we essentially see the same behaviour in 2005 and 2010. Since the scope issue is not the focus of this paper and since there were no statistically significant differences between the two split-samples when comparing socio-demographic variables nor when comparing WTP estimates, and also for reasons of simplicity and statistical efficiency, we have chosen to pool the data from the two split-samples for the analyses in this paper.

attributed to either the social dimension (preference heterogeneity) or the spatial/commodity dimension (differences in sites or the service being valued, e.g. fishing at different sites or fishing and hunting at the same site), but only to the temporal dimension, i.e. the pure time effect of the 5 years gone by. We refer to this sample of respondents who validly participated in both surveys as the test–retest sample. In addition to the transfer in the test–retest sample, temporal BT is also assessed by transferring valuation results from all 2005 respondents to the new respondents from 2010. This is equivalent to what has been done in the majority of previous temporal BT studies, where the transfer represents simultaneously transferring values across time and across different individuals. We refer to this sample as the non-identical sample.

3.2 Methods Used to Obtain Values for Transfer

To assess transferability of unit values, we simply transfer the observed non-parametric mean WTP values and use the associated variances for testing. To assess transferability of benefit functions, however, we need to estimate bid functions. The type and format of survey guides the model used to estimate WTP. As this study uses open-ended CVM, the respondent is expected to state the maximum amount he/she is willing to pay for the proposed flood risk reduction project. An initial visual inspection of the bids revealed that they range from zero to 10,000 DKK (1 DKK ∼ €0.13) with a substantive point mass at zero for all sample groups. An appropriate model for such data is the Tobit model (Verbeek 2008; Wooldridge 2010). Following Verbeek (2008) and Wooldridge (2009, 2010), the standard Tobit model in latent form is:

$$y = \max\left(0, y^*\right), \quad y^* = x\beta + u \tag{5}$$

where u is assumed normally distributed with mean zero and variance σ^2, y is the observed WTP, y^* is the latent variable indicating the WTP for each respondent as determined by the model variables, while x and β are vectors of explanatory variables and population parameters respectively. Since the Tobit model as such is not the main focus here, we do not provide a detailed theoretical account of the rather standard econometric model here. Instead, the reader is referred to for example Carlsson et al. (2012) for a thorough econometric specification of an applied Tobit model.

3.3 Tests to Assess Transferability

Overall, two approaches have previously been used to evaluate BT transferability. The most commonly used method is to assess the size of transfer error (TE). This error shows the percentage difference between the actual and the transferred value and can be written as:

$$Transfer\ Error\ (TE) = \frac{(Transferred\ value - Actual\ value)}{Actual\ value} * 100 \tag{6}$$

The lower the size of TE, the better is the transfer performance (Bergland et al. 1995). It is evident from Eq. (6) that if the transferred value equals the actual value then the size of the TE will be zero. The size of the TE may be deemed important to assess the net savings associated with conducting a BT instead of a primary valuation survey; the saved costs that otherwise would have been used to conduct a primary survey should be considered relative to the added uncertainty as reflected by the TE. However, an original study is required at the policy site to obtain TEs, implying that decision makers using BT cannot know the actual size of the TE, though they might be able to predict it from previous experience. Moreover, since the desired degree of precision varies across disciplines, purposes and contexts, there is no standardised acceptable size of TE (Allen and Loomis 2008; Colombo and Hanley 2008; Johnston and Rosenberger 2009; Baskaran et al. 2010). Apart from these practical limitations, TE serves an important purpose in BT transferability studies where reductions or minimisation of the TE measure corresponds to increased reliability and validity of BT.

The second approach used to evaluate transferability covers statistical tests for equality and/or equivalence of the transferred and actual values (Bergland et al. 1995; Brouwer and Spaninks 1999; Baskaran et al. 2010). These tests include both parametric (which assume a specific distribution) and non-parametric (without distributional assumptions) approaches. Recommended tests of BT are summarised in Table 2 following Bergland et al. (1995) and Brouwer and Spaninks (1999).

The *t* test and Mann–Whitney tests (Mann and Whitney 1947) for unit value BT have been used to test if two samples have statistically identical mean WTP values. The choice between these two tests depends on the distribution of the WTP values (Brouwer and Spaninks 1999). These tests have also been used to test transferability of benefits estimated from bid functions. The transferred WTP value (computed using a transferred equation and mean values of covariates from the policy site) is tested against the actual benefit (WTP) obtained from the policy site with its own estimated bid function and mean values of covariates.

The simple likelihood ratio (LR) test for testing the validity of BT is conducted by pooling the two samples together and incorporating a dummy variable for one of the samples in the bid function. The LR test statistic is two times the difference between the value of the maximum likelihood function for the unrestricted function (including the dummy variable) and the restricted function (excluding the dummy variable). Under the null hypothesis of an insignificant coefficient of the dummy variable, the LR test statistic follows a chi-squared distribution with one degree of freedom. Rejection of the null hypothesis is considered as non-transferability of the benefit function (Brouwer and Spaninks 1999; Brouwer and Bateman 2005).

The Chow test is developed to test coefficient stability of two (or more) linear models estimated by linear least squares methods. Its applicability to a Tobit model, which is estimated by maximum likelihood, is however not yet clearly documented in the literature. Notwithstanding this, Bergland et al. (1995) used a Chow test version of the LR test. The test requires conducting three regressions: an independent regression for each group and one for a pooled sample. The test statistic for the present study would be:

$$LR_{calculated} = -2 \times (ln\ L_{2005+2010} - (ln\ L_{2005} + ln\ L_{2010})) \tag{7}$$

Where $L_{2005+2010}$ is the maximum value of the likelihood function from the pooled dataset, L_{2005} and L_{2010} are maximum values of the likelihood functions obtained from the estimation of the 2005 and 2010 surveys respectively. Under the null hypothesis of identical distributions, the LR test statistic follows a chi-squared distribution.

The Wald test is the other transferability test most often recommended for assessing benefit function transferability (Brouwer and Bateman 2005; Brouwer 2006). Here, the equality of estimated coefficients is tested. It is preferred to the tests discussed above, since other tests deal with statistical equality of the distributions rather than equality of coefficients. Specifically, the abovementioned tests do not necessarily test for equality of the study and policy site coefficients which are transferred in the benefit function transfer (Bergland et al. 1995).

Table 2 Statistical tests of transferability

Transfer type	Statistical test	Null hypotheses[a]	Brief description of the test
Unit value transferability tests	*t* test	$\overline{WTP}_s = \overline{WTP}_p$	Assumes normal distribution and tests if the mean WTP value at the policy site ($\overline{WTP}_p$) equals mean WTP at the study site ($\overline{WTP}_s$)
	Mann–Whitney test		Non-parametric test that combines the ranks of the two distributions and tests for rank sum differences between the samples
Benefit function transferability tests	Likelihood ratio (LR) test	$\beta_{d1} = 0$	From pooled data regression, tests if a dummy for survey year is significant. Under the null hypothesis of an insignificant survey year dummy, the study and policy sites' distributions are said to be statistically identical, and thus transferable
	Chow test version of LR test	$\check{\beta}_s = \check{\beta}_p = \check{\beta}$	Tests if individual functions of the study and policy sites differ from a common benefit function. Under the null hypothesis the study and policy site distributions are drawn from an identical population distribution
	Wald/score test	$\beta_s = \beta_p$	Tests if the study and policy sites' benefit functions have statistically identical coefficients
	t test	$\overline{f\left(X_p, \beta_s\right)} = \overline{f\left(X_p, \beta_p\right)}$	Tests if the transferred and actual benefits calculated from benefit functions are statistically identical
	Mann–Whitney test	$\overline{f\left(X_p, \beta_s\right)} = \overline{f\left(X_p, \beta_p\right)}$	Tests if the transferred and actual benefits calculated from benefit functions are statistically identical (non-parametrically)

[a] *s* denotes study site and *p* denotes policy site while d1 denotes a dummy variable for either policy or study site

4 Results and Discussion

4.1 Descriptive Results: Comparison of Study and Policy Sites

Both the study and policy sites refer to the same geographical site, namely an area along the western coast of Jutland, Denmark. The physical nature of the site was the same in both surveys since there was no recognised change in sea level nor were there any flooding events between the two survey years. Therefore, any observed valuation differences between the two surveys are unlikely to be attributed to physical changes to the site.

To test for potential differences in demographic characteristics and attitudes across samples, Table 3 provides an overview of the central tendencies in selected demographic variables as well as attitudinal variables acquired from the corresponding questions in the questionnaire. Chi-square tests for equality are conducted for the two different groups of samples. The first group is the test–retest sample while the second is the non-identical sample.

Table 3 reveals that many of the variables of interest are not statistically different across the two survey years for the test–retest sample. For example, the percentage of respondents whose yearly household income is less than 300,000 DKK (Income_300) remains fairly constant over time. During the period, five respondents took some form of higher education (percentage share increased from 24 to 27 %) while 10 respondents lost their permanent employment (percentage share decreased from 65 to 58 %), but neither of these differences were statistically significant. Other minor changes for the test–retest sample include attitudinal variables such as the perception of expected location above sea level (Expect_dist), perception of the risk of flooding (Flood_risk_safe) and previous flooding experience (Experience). Again, these differences were not significantly different at the 5 % level of significance. The statistically significant differences for the test–retest group include age (which is not surprising since every respondent should be 5 years older) and two attitudinal variables: perception of the degree of global warming having an effect on flooding (Global_warming) and believing that there is a need for better flood protection (Should_protected).

As expected, there are more differences among the respondents in the non-identical samples compared to those in the test–retest samples. Compared to the sample with all of the 2005 respondents, the new respondents from 2010 are relatively wealthier, older, have a more equal gender distribution, perceive themselves to reside in a safer area, and believe that they need better flood protection. However, most of these differences are not statistically significant. The only statistically significant differences at a 5 % level of significance include the variables representing the age of the respondents (Age), residence location (Sample1), and two attitudinal variables: Should_protected and Global_warming. Education, Income_300 and Income_800 are also significantly different for these non-identical groups of respondents, though only at a 10 % level of significance.

Overall, it seems fairly evident that in both the test–retest and non-identical samples there are increased expectations of future flooding events as a consequence of increased global warming. This is not surprising considering the generally increasing public attention paid to the global warming issue during the period of 2005–2010, particularly in Denmark where the highly profiled United Nations Climate Change Conference COP15 took place in Copenhagen in 2009.

4.2 Tobit Model Results

A Tobit model is used to estimate bid functions for each of the four sample groups. All estimates and corresponding predicted values are obtained from statistical best-fit models.

Table 3 Demographic and attitudinal equality tests

Variable	Description	Percentage share/average				Chi-square equality test (p values in parentheses)	
		Test–retest 2005	Test–retest 2010	All 2005	New 2010	Test–retest 2005 versus 2010	All 2005 versus new 2010
N	Sample size	150	150	581	247	–	–
Income_300	Dummy for yearly income below 300,000 DKK	25 %	25 %	29 %	23 %	0.0011 (0.97)	2.83* (0.092)
Income_800	Dummy for yearly income between 700,000–800,000 DKK	5 %	10 %	14 %	1 %	2.31 (0.13)	3.22* (0.073)
Male	Dummy for male	63 %	63 %	55 %	51 %	0.014 (0.91)	1.39 (0.24)
Age	Age of respondent	50	55	52	55	3.8*** (<0.001)	2.57** (0.01)
Job	Dummy for full-time work or self-employment	65 %	58 %	59 %	54 %	1.41 (0.24)	1.55 (0.21)
Education	Dummy for having higher education	24%	27 %	21%	27 %	0.4368 (0.51)	3.2661* (0.071)
Sample1	Dummy for residing in most flood exposed location	15 %	15 %	19 %	4 %	0.0000 (1.00)	29.61*** (<0.001)
Responsibility	Dummy for attitude of feeling individually responsible for paying for better protection	50 %	53 %	52 %	51 %	0.21 (0.64)	0.027 (0.87)
Should_protected	Dummy for attitude of believing the area needs better protection	11 %	56 %	12 %	59 %	69.36*** (<0.001)	205.33*** (<0.001)
Global_warming	Dummy for perceiving that flooding risk will increase due to global warming	43 %	65 %	43 %	73 %	12.88*** (<0.001)	63.55*** (<0.001)
Experience	Dummy for experiencing flood	58 %	51 %	52 %	46 %	1.63 (0.2)	2.08 (0.15)
Flood_risk_safe	Dummy for not feeling at risk of flood	79 %	77 %	76 %	75 %	0.18 (0.67)	0.056 (0.81)
Expect_dist	Dummy for perceiving that residence is located <5 m above sea level	69 %	64 %	61 %	62 %	0.73 (0.39)	0.052 (0.82)

*, ** and *** denote 10 %, 5 % and 1 % levels of significance respectively. Chi-square tests are conducted on the basis of the actual numbers behind the percentages

Table 4 Summary of Tobit model regression results for the different sample groups

Variable	Coefficient			
	Test–retest 2005	Test–retest 2010	All 2005	New 2010
Income_300	−506.23* (281.92)	−286.30* (153.34)	−404.56*** (148.77)	
Income_800				1,074*** (387.5)
Male			34.34*** (122.29)	
Age			−8.49* (4.78)	
Job				418.1** (200)
Education	−486.82* (287)	−328.59** (148.93)		
Sample1	794.85** (333.69)	489.66*** (179.44)		
Responsibility	439.22* (246.88)	218.7* (130.69)	256.56** (123.05)	
Should_protected	−794.17** (403.96)			
Global_warming			383.26*** (124.76)	
Flood_risk_safe			−219.90 (143.1)	
Constant	736.94*** (211.07)	713.97*** (113.71)	799.51*** (305.32)	335** (149.7)
Log (scale)	7.23*** (0.0066)	6.67*** (0.061)	7.24*** (0.034)	7.33*** (0.05)
Sample size	150	149[a]	581	247
Overall model significance	$\chi^2_{(5)} = 16.3$***	$\chi^2_{(5)} = 18.37$***	$\chi^2_{(6)} = 46.13$***	$\chi^2_{(2)} = 12.7$***

*, ** and *** denote 10 %, 5 % and 1 % levels of significance. Standard errors are in parenthesis.
[a] Due to a missing observation for one of the explanatory variables, the Test–retest 2010 model has 149 observations instead of 150

All estimates of covariates in the models presented in Table 4 are significantly different from zero at a 10 % level of statistical significance or lower. Table 4 reveals that the covariates explaining the WTP of respondents vary across sample groups over time, though not within the test–retest sample. More specifically and with the exception of one attitudinal covariate (Should_protected), the determinants of WTP remain identical over time for the test–retest sample. That is, factors determining WTP remain fairly stable from 2005 to 2010 when considering only the pure temporal effects. However, the determinants of WTP differ when the comparison is both over time and across different respondents. This would suggest that temporal benefit function transfer over the 5-year period is reasonably valid when transferring only over time, but when also transferring between respondents, the transfer of a benefit function may be questionable or at least less valid.

4.3 Transfer Errors

In the following we conduct the empirical temporal transfer of values in terms of unadjusted, income adjusted and benefit function transfers for the two groups (test–retest and non-identical). Table 5 displays the obtained TEs.

Table 5 Benefit transfers and transfer errors

Transfer type	Mean WTP (DKK)	Test–retest sample	Non-identical sample	Average TE
Unadjusted unit value transfer	Actual 2010 value	779	795	
	Transferred value	900	770	
	Transfer error	15.5 %	−3.0 %	9.3 %
Income adjusted unit value transfer	Actual 2010 value	779	795	
	Transferred value	1,017	870	
	Transfer error	30.5 %	9.5 %	20.0 %
Benefit function transfer	Actual 2010 value	816	979	
	Transferred value	781	834	
	Transfer error	−4.4 %	−14.8 %	9.6 %

Compared to previous BT studies, the average TE of 9.3 % for the unadjusted unit value transfer is at the low end. This would indicate that values are relatively stable and, thus, for most purposes still transferable after 5 years. The test–retest sample transfer overestimates the 2010 value by 15.5 % which is in line with previous similar studies that find declining WTP over time. Slightly surprising though, is that with a TE of only 3 %, the transfer for the non-identical sample underestimates the 2010 value. The relatively better performance of this transfer is unexpected[8] since it transfers over two dimensions (temporal and social) with two potential sources of error as opposed to only one source of error in the test–retest sample. This would suggest that the 15.5% reduction in WTP caused by the pure time effect is by-and-large counteracted by an increase in WTP caused by differences between the respondents in 2005 and the new sample of respondents in 2010.

Theoretically, the unadjusted unit value transfer becomes less and less appropriate the larger the differences between the study and policy sites. One possible transfer improvement in such cases is to adjust for observed differences. Bateman et al. (2002) categorised adjustment factors into differences in socio-economic and demographic characteristics of the relevant populations, physical characteristics, the proposed change and market conditions applying to the sites. Many of the these factors that could be considered when conducting adjustments to unit value transfers are not relevant for this study since the site is the same and does not have any physical change. Moreover, the sampling frame is also the same for both surveys. For the non-identical samples there are, as displayed in Table 3, some differences in demographics, and in particular a decrease in disposable income in 2010 compared to 2005. Even though the respondents in our test–retest sample are identical, there are also some differences in this sample in the level of disposable income of respondents in the 2 years. According to economic theory, income should have a significant bearing on WTP—as confirmed by the Tobit model shown in Table 4. Hence, it would seem relevant to conduct an income adjusted unit value transfer to account for these differences in income.

The disposable income of an average Danish household in the survey area of interest was 296,993 and 336,641 DKK in 2005 and 2010, respectively (Statistics Denmark 2012). The income adjusted unit value transfer presented in Table 5 is computed using these mean incomes. Contrary to expectations, Table 5 reveals that the income adjusted unit value transfer performs worse than the unadjusted unit value transfer. In fact, the average TE is more than double (20 vs. 9.3 %). There is also a relatively poorer performance of the income-adjusted

8 Recent findings, however, suggest that different populations, locations and even commodities can produce similar WTP distributions (Moeltner and Rosenberger 2012; Johnston and Moeltner 2013).

transfer for the test–retest sample when comparing the two transfers. Here, the TE increases from 15.5 to 30.5 % after adjusting for income differences.

Moving to the performance of benefit function transfers, the lower third of Table 5 shows the transferred values and the actual values computed using the regression results and corresponding mean values of explanatory variables obtained from Table 4. Recall that transferred values are obtained by inserting the mean values of explanatory variables from 2010 into the transferred benefit function from 2005.[9] The overall performance, relative to previous studies testing function transfers, is fairly decent with an average TE of 9.6 %. Interestingly, the benefit function transfer performs better than the unadjusted and income adjusted unit value transfers for the test–retest sample transfer. Table 5 reveals that it is the benefit function transfer that obtains the lowest TE (4.4 %) for the purely temporal transfer over identical respondents. Comparing transfer performance of unadjusted unit value transfer and benefit function transfer reveals another interesting result; while unadjusted unit value transfer performs better for the non-identical samples, benefit function transfer is better for the test–retest sample. This would suggest that the function transfer works best when transferring only over time but not when also transferring across different respondents. As stated previously, unit value transfers may perform better for transfers across similar sites since they, unlike function transfers, do not over-parameterise transfers between similar sites. Function transfers, however, are generally preferred to unit value transfers for dissimilar sites as they increase accuracy by adjusting for factors such as social and physical differences between study and policy sites (Bateman et al. 2011). For our test–retest sample, the experimental design ensures that there are virtually no social or physical differences. In line with Bateman et al. (2011), this would suggest that a simple unit value transfer should perform as well as or even better than a function transfer for our test–retest sample while the opposite is likely to be the case for the transfer between non-identical respondents. However, the better performance of the function transfer relative to the unit value transfer in our test–retest sample should not necessarily be considered as counterintuitive to the above expectation, given the observed changes over time in age and some attitudinal variables as shown in Table 3. Similarly, the poor performance of the function transfer relative to the unit value transfer for the non-identical respondents could be due to the two samples not having similar covariates as indicated by Table 4.

A related issue concerns which explanatory variables to include in the transferred bid functions. Whitehead and Hoban (1999) found that statistical best-fit functions including generic, theory-derived as well as ad-hoc attitudinal variables were transferable, but transferability reduced markedly when attitudinal variables were not included. Similarly, Zandersen et al. (2007) and Berrens et al. (2000) suggested the inclusion of a wide range of covariates to improve the performance of BT. However, other studies found that the benefit function is transferable only when theory-derived covariates are considered and not when statistical best-fit functions including ad-hoc contextual and attitudinal variables are used (Brouwer and Bateman 2005; Brouwer 2006; Bateman et al. 2011). We have also conducted such a function transfer approach by considering income as the only theory-derived covariate to enter

[9] There is some discrepancy between studies concerning the computation of TE from benefit functions, as also noted by Bateman et al. (2011). Some studies compare transferred predicted values with the simple mean at the policy site (e.g. Brouwer and Bateman 2005; Bateman et al. 2011), while others compare transferred predicted values with the predicted value at the policy site (e.g. Barton 2002; Rozan 2004; Eshet et al. 2007). We follow the latter approach, since it reasonably compares two corresponding values; otherwise we could have additional TE from comparing the simple mean and the predicted value of the same study site (e.g. this value equals 4.7 % in our 2010 test–retest sample) as estimation functions do not explain 100 % of the actual value.

the function. The resulting TEs were 26.2 and 14.0 % for the test–retest and non-identical samples respectively. The TE for the test–retest sample is more than five times as large as the TE for the statistically best-fit function. This finding is in line with Zandersen et al. (2007) and Whitehead and Hoban (1999) but somewhat at odds with Bateman et al. (2011). For the non-identical samples, however, the theory-derived covariate approach does slightly improve the TE (from 14.8 to 14.0 %). These results do not change much if we also include residence location (Sample1) as a theory-derived variable.[10] Here, the TE from the test–retest sample (26.3 %) is almost identical to when income was the only included variable, while there is a further slight improvement for the non-identical samples (13.0 %). This suggests that TE is not markedly improved in our case when we exclude ad-hoc contextual variables and rely solely on theoretically derived covariates. For the test–retest sample the TE is actually considerably worsened.

Another concern is also whether these theory-derived, contextual and attitudinal ad-hoc variables are even available at the policy site, since practical applicability of function transfer requires the availability of at least central tendencies for all explanatory variables at the policy site. Function transfers applied in BT studies implicitly assume that all necessary data are available at the policy site from existing reliable secondary data sources for all explanatory variables used in the transferred benefit function. However, in our case it would not be possible to obtain data on for example the attitudes and flood risk perceptions from available data sources. Central tendencies would only be available for socio-demographic variables like age, income, gender, education, residence location and employment. We have tested a function transfer based on bid functions using only these readily available data, to assess how large the TE would have been if BT had been used in a practical BT setting to elicit WTP in 2010 instead of the primary study. The TEs obtained in this more practically oriented approach to function transfer are 33 and 97 % for the test–retest samples and the non-identical samples respectively. This is more than six times the transfer errors found in the function transfer based on statistical best-fit functions in Table 5. Even though the relatively higher TEs are generally expected in practical applications of BT, since data may not be available for the entire relevant set of covariates, our considerably larger TEs are not expected and may cast doubt on the reliability of BT in situations where there is quite limited data at the policy site. This could suggest that the TE in practical applications, where there is limited secondary data, could be higher than the TE from studies assessing the reliability of BT found from primary data. BT studies could therefore be somewhat at risk of providing an overly optimistic impression of the performance of function transfers in common practice.

4.4 Transferability Tests

In this section, the statistical tests outlined in Table 2 are conducted. The *t* tests and Mann–Whitney (M–W) tests displayed in Table 6 are used to assess transferability of unit values. The null hypothesis of equality of observed and transferred WTP values cannot be rejected by the M–W test for the unadjusted and the income adjusted unit value transfers. The result is the same for the *t* test, except that the income-adjusted transfer for the test–retest sample has a rejection of the null hypothesis of transferability at the 5 % level of significance. One explanation for this rejection could be because of the skewing effect of the income adjustment. That is, the income adjustment does not consider the possibility of a temporal WTP shift from zero to positive values. In doing so, the income adjustment increases the variation in WTP,

[10] Residence location can be included as a theory-derived variable, since it may be argued based on economic theory that respondents residing in the area with the highest risk of flooding would also have a higher WTP to reduce the risk, assuming that there is no difference in risk aversion between these respondents and others.

whereas the real effect of income at the policy site could include shifts from zero to positive WTP. This could also explain why we observe a higher TE for the income-adjusted transfer than the unadjusted unit value transfer in Table 5. The M–W test is less likely to be affected by the income adjustment, since the test considers the sum of the ranks, not the means.

Table 4 revealed that the benefit function in 2005 is similar to the benefit function in 2010 for the test–retest sample. In terms of the number of explanatory variables, the two equations differ only by one attitudinal variable (Should_protected), in that this variable is statistically significant only in 2005. However, the benefit functions for the non-identical samples are quite different for the 2 years. This is interesting, in that the similarity of benefit functions for the test–retest sample, which is not seen for the other samples, could help to explain why several previous studies found non-transferability of statistical best-fit functions. It is possible that the best-fit functions could have been transferable if transfers over identical respondents had been done, i.e. considering the pure temporal transfer. Table 7 shows the results of the statistical tests of transferability of benefit functions.

The simple LR tests cannot reject that the distributions of WTP values are identical over time, both for the test–retest sample and the non-identical samples. A similar result is revealed by the Wald test in which the null hypotheses of statistical equality of coefficients of the study and policy sites cannot be rejected at conventional levels of significance. Unlike the results reported by Downing and Ozuna (1996), Table 7 also reveals that benefits obtained from benefit functions are statistically transferable (see t test and M–W test results). Transferability of benefit functions for the non-identical samples is unexpected given that the two benefit functions have different covariates as we saw in Table 4. We expected the Wald test to reject equality of coefficients for at least the Age and Global_warming variables, since these are significant only in the 2005 benefit function and also have significant differences between samples as seen in Table 3. The equality of coefficients for the covariates income, gender and responsibility may not be surprising since these variables are not statistically different among the non-identical samples.

However, the result is different in the case of the Chow test version of the LR test. Here, statistical equality of the pooled data distribution and the two individual dataset distributions is rejected for the test–retest sample (and at 10 % level of significance for the non-identical samples). This test result is unexpected for the test–retest sample since the two benefit functions differ only by one explanatory variable and they have statistically equal WTP distributions. Nevertheless, this test does not noticeably alter our overall conclusion of transferability of values for the test–retest sample since the test does not concern itself either with equality of WTP values or equality of estimated coefficients.

4.5 Explaining Temporal Changes in WTP

One source of TE could be differences between the policy and study sites. In relation to the pure temporal dimension of transferability, another central source of TE could be changes in preferences over time. Due to the fact that we have a test–retest sample where the only transfer dimension is time, we have a unique opportunity to assess whether the observed transfer errors can be ascribed to changes in preferences. For each of the 150 respondents in the test–retest sample, we know whether their WTP has increased, decreased or remained constant from the 2005 to the 2010 survey. Table 8 provides an overview of changes observed for the individual respondents.

As Table 8 reveals, about 39 % of respondents stated the same WTP in 2010 as in 2005, while 24 % reduced their stated WTP and 37 % increased it, with a net decline in mean WTP of about 121 DKK. This decline in WTP came from 36 respondents who reduced their

Table 6 Mann–Whitney (M–W) and *t* test results for unit value transfers

Transfer type	Sample	Mean WTP equality test (*p* values in parenthesis)		Transferability at 5 % level of significance
		M–W test	*t* test	
Unadjusted unit value transfer	Test–retest 2005 versus test–retest 2010	$z_{(298)} = -1.11$ (0.27)	$t_{(298)} = 0.96$ (0.34)	Transferable
	All 2005 versus new 2010	$z_{(826)} = -0.75$ (0.45)	$t_{(826)} = -0.26$ (0.8)	Transferable
Income-adjusted unit value transfer	Test–retest 2005 versus test–retest 2010	$z_{(298)} = 1.42$ (0.16)	$t_{(298)} = 2.074$ (0.039)	Transferable by M–W test but not by *t* test
	All 2005 versus new 2010	$z_{(826)} = 1.95$ (0.052)	$t_{(826)} = 1.18$ (0.24)	Transferable (but barely so by M–W test)

Table 7 Statistical tests of transferability of benefit functions

Test	Test–retest 2005 versus test–retest 2010		All 2005 versus new 2010		Transferability at 5 % level of significance
	Test statistic	p value	Test statistic	p value	
Simple LR test	$\chi^2_{(1)} =< 0.001$	0.99	$\chi^2_{(1)} = 0.04$	0.84	Transferable
Chow test version LR test	$\chi^2_{(7)} = 49.0$	< 0.001	$\chi^2_{(9)} = 15.53$	0.077	Transferable though not for the test–retest sample
Wald test	$\chi^2_{(5)} = 8.8$	0.12	$\chi^2_{(6)} = 5.47$	0.49	Transferable
Mann–Whitney test	W $= 12{,}555$	0.05	W $= 29{,}620$	0. 86	Transferable (though barely so for the test–retest sample)
t test	$t = 0.15$	0. 88	$t = 1.47$	0. 14	Transferable

Table 8 Summary of temporal development in individuals' stated WTP from 2005 to 2010, in DKK

Individual change in WTP	No. of respondents (% of total)	WTP bid in 2005		WTP bid in 2010 (% change)	
		Median	Mean	Median	Mean
Increasing WTP	56 (37 %)	200	363	700 (250 %)	1,011 (179 %)
Decreasing WTP	36 (24 %)	1,200	2,260	350 (−71 %)	746 (−67 %)
No change (zero bid)	7 (5 %)	0	0	0	0
No change (positive bid)	51 (34 %)	500	653	500	653
Total test–retest sample	150	500	900	500 (0 %)	779 (−13 %)

Table 9 Linear model estimates explaining WTP differences

Variable	Coefficient
Increase in income	248.94* (127.50)
Sample1	315.65** (151.26)
Decrease in demand for flood protection	−278.27*** (99.78)
WTP stated in 2005	−0.72*** (0.06)
Constant	451.72*** (69.13)
Sample size	150
R-squared	0.67
F(4, 145)	37.4

*, ** and *** denote 10 %, 5 % and 1 % levels of significance. Standard errors are in parenthesis

WTP by an average of 1,514 DKK. Even though more respondents increased their WTP than decreased it, the average increase was somewhat lower at an average of 648 DKK, i.e. not high enough to counter the average reduction.

It is evident from the mean and median values that, on average, respondents stating a relatively low bid in 2005 are likely to more or less double their bid in 2010, whereas a relatively high bid in 2005 is likely to be more than halved in 2010. In between the very high and low bids, respondents tend to stick to their original WTP bid from 2005. The question is whether the changes in an individual's WTP can be ascribed to other characteristics of the individual, either socio-demographic or changes in attitudes or behaviour. We report results from two models:[11] a linear model investigating explanations for the observed changes of WTP (Table 9) and a multinomial logit model exploring factors that induced respondents to increase, decrease or not to change their WTP (Table 10).

The dependent variable in Table 9 is the WTP difference between 2005 and 2010 obtained from the test–retest respondents. The explanatory variables could include all those listed in Table 3. However, we found only four variables that significantly explain the WTP differences. The linear regression model reported in Table 9 shows a reasonable fit to the data and the variables have the expected signs. For instance, the variable representing the WTP stated in 2005 has a negative sign, implying that WTP in 2010 decreased by about 0.72 DKK, on average, for each DKK that the respondent was willing to pay in 2005. As noted previously,

[11] A model to test for carry-over effects was also conducted, which showed that a dummy variable representing participation in both survey years was not significant. In other words respondents do not seem to be drawing on their experience of answering the survey in 2005 when answering in 2010.

Table 10 Multinomial Logit model odds-ratio estimates explaining WTP change over time

Explanatory variables	WTP decrease	WTP increase
WTP stated in 2005	0.0017*** (0.0001)	−0.0014 * * * (0.0001)
Age	−0.051*** (0. 0025)	−0.0123*** (0.003)
Change in should_protect (dummy = 1 if change, 0 otherwise)	0.87*** (0. 053)	0.83*** (0.07)
Change in global_warming (dummy = 1 if change, 0 otherwise)	2.3*** (0.102)	2.23*** (0.17)
Change in flood_risk (dummy = 1 if change, 0 otherwise)	−1.3*** (0.085)	0.56*** (0.17)
Constant	−0.034 (1.43)	0.74 (1.035)
Number of observations	150	
Wald chi^2(10)	40.31	
Pseudo-R ^{2}	0.28	

*, ** and *** denote 10 %, 5 % and 1 % levels of significance. Standard errors are in parenthesis

this decline in WTP is not uncommon, see for example Whitehead and Hoban (1999) and Brouwer and Bateman (2005). The attitudinal variable representing respondents no longer seeking flood protection contributes negatively to the change in WTP, as could be expected. On the other hand, residing in relatively riskier areas (Sample1) and an increase in income both contribute positively to the WTP change.

We now turn to the multinomial logit model results reported in Table 10 where the dependent variable is an unordered response with three outcomes relating to the change in WTP: increase, decrease and no change. The explanatory variables are WTP stated in 2005, the age of the respondent and three variables representing a change in either direction of attitudinal variables. The estimation results in terms of the relative probability ratio (log-odds-ratios)[12] between two response outcomes are presented in Table 10, where the base outcome is no change. The values in the parentheses are standard errors obtained by the delta method.

When a respondent is willing to pay one more DKK in 2005, this increases the log-odds ratio between decreasing and not changing WTP in 2010. In other words, the higher the bid in 2005, the greater the probability of stating a lower bid than the same bid in 2010. For a WTP increase, the sign of the log-odds ratio is negative indicating that the higher the bid in 2005, the lower the chance of stating an even higher bid in 2010 than stating the same as in 2005. Overall, this could reflect respondent imprecision or uncertainty. Rather than having perfect knowledge of own maximum WTP in the somewhat unfamiliar situation that a CVM questionnaire poses to most people, respondents may only know an approximate interval in which their maximum WTP lies (e.g. Dubourg et al. 1994; Ready et al. 2001; Håkansson 2008). So if a respondent by chance reported a relatively high bid from this range of maximum WTP in 2005, there is a higher probability of reporting a lower or the same WTP in 2010 than reporting a higher WTP (assuming a normal distribution of the WTP interval). The signs for the log-odds ratios for the age of the respondent indicate that the older the respondent, the less likely they will be to change their WTP, either by increasing or decreasing their

[12] As Wooldridge (2010) illustrates, the coefficients of the variables do not have a direct interpretation, and even the sign of coefficients may not necessarily indicate the direction of the effect. One way of interpreting the results is to compute the log-odds, which is the relative probability ratio between two response outcomes (see Wooldridge (2010) for the computation).

original bid. In other words, older respondents would seem to exhibit more stable WTP bids. As expected, the remaining log-odds ratios for changes in the attitudinal variables show that changes in these attitudes have significant impacts on the likelihood of respondents changing their WTP.

From Tables 9 and 10, it is evident that the change in WTP which was reflected in the TE is explained, at least partially, by the age of respondents, the amount of WTP respondents were willing to pay in 2005 and the change in attitude of respondents. Residence location and change in income also contribute to the observed difference in WTP, though they are only statistically significant in the linear model. All of these explanatory variables have the expected signs, and as such it serves as a survey validation in terms of indicating that the changes we see in WTP in the temporal dimension are behaviourally meaningful and not just more or less random artefacts of the survey setup.

5 Conclusion

In this study, we find that open-ended CVM valuation results concerning WTP for flood risk reductions are temporally transferable over a time horizon of 5 years, specifically from 2005 to 2010. Depending on the type of transfer, we find TEs ranging from 3 to 97 %. Being amongst the lowest TEs observed in the BT literature, the lower bound here would surely be acceptable for almost any practical purpose. However, whether the upper bound at almost 100 % could be considered an acceptable TE or not, would depend on the required degree of precision which varies across disciplines, contexts and transfer purposes (Allen and Loomis 2008; Colombo and Hanley 2008; Johnston and Rosenberger 2009; Baskaran et al. 2010). If for instance the 2005 value estimates were to be transferred and used as benefit estimates in a Cost-Benefit Analysis in 2010, a transfer error of 100 % may be considered acceptable if the benefit-cost ratio is above 2.

When comparing BT approaches of unadjusted unit value transfer and benefit function transfer based on a statistical best-fit model, we see them perform equally well with TEs just below 10 % on average. The income adjusted unit value transfer performs somewhat worse with a 20 % TE on average. Relating to the rather ambiguous results in the existing literature concerning the relative performance of different types of BT, and particularly whether function transfer is to be preferred over the simple unit value transfer (see e.g. Bateman et al. 2011 and Kaul et al. 2013), our results as such do not offer further conclusive evidence. Our experimental setup allows us to separate the purely temporal element of TE from the social and spatial elements. Using a transfer over identical respondents in a test–retest setting and thereby ensuring that any TE is entirely related to a pure effect of time, we find that the benefit function transfer with a TE of only 4.4 % outperforms the unadjusted unit value transfer (TE = 15.5 %) and income adjusted unit value transfer (TE = 30.5 %). While this would speak in favour of using function transfer, neither economic theory or the existing literature seems to support the notion that function transfer should perform best in this particular setting where the only transfer dimension is across time. Benefit function transfers have mainly been argued to perform relatively better for transfers between dissimilar sites and heterogeneous populations, i.e. where the social and spatial dimensions are particularly important (Chattopadhyay 2003; Brouwer and Bateman 2005; Bateman et al. 2011), which is clearly not the case in our survey. Furthermore, when allowing for both temporal and social elements in TE by transferring across different respondents, the simple unadjusted unit value transfer turns out to perform best with a TE of only 3.0 %, followed by income adjusted unit value transfer and benefit function transfer with TEs of 9.5 and 14.8 %, respectively. Since this is a more relevant setting for most practical BT purposes, our results would favour the use of simple unadjusted unit value transfer. It should be noted, though, that most practical BTs would also entail a spatial element in the TE. Since this dimension is kept constant in our experimental setup, further research is needed in order to ascertain the impact of this specific dimension on the relative performance of the different transfer methods.

A range of statistical tests support that our 2005 value estimates are transferable to 2010, for the pure temporal transfer as well as when confounding the temporal and the social dimension. Applying *t* tests and Mann–Whitney tests, both unit value transfers are found to be statistically transferable, though the results are ambiguous for the income adjusted unit value transfer. For the statistically best-fit benefit function transfer, the functions are found to be largely transferable based on five different statistical tests, both when transferring across different respondents and across identical respondents. However, for the transfer across different respondents the statistical equality of the estimated benefit functions for the 2005 and 2010 surveys is unexpected given that these two best-fit benefit functions had different significant covariates in the two survey years. This might indicate that the statistical tests are not particularly rigorous and the transferability of the benefit function across different respondents should be viewed with some caution. Furthermore, for the transfer across identical respondents the results are slightly ambiguous since one of the five tests significantly rejects transferability of the benefit function and another is borderline significant. Nevertheless, the overall picture would suggest that our functions are indeed transferable.

Importantly we find that BT tests may lead to potentially false confidence in the precision of BT under practical BT data constraints. A potential limitation in many BT function transfer tests is that they rely on all available data; also that which is only available from a primary valuation survey at the policy site (or in our case in the policy year). Considering that the main purpose of doing BT is typically to avoid an expensive and time consuming primary valuation survey, such data is typically not available in practice. We therefore also conducted a more practically oriented function transfer, only making use of secondary data. This function transfer resulted in substantively higher TEs, which may not be surprising, since some variables that could explain WTP, but cannot be obtained without conducting a primary survey at the policy site, are unavailable and thus excluded when doing practical BT. Nevertheless, this result underlines that one should exert caution when testing benefit function transfer, since the use of policy site primary data in benefit functions may reduce TEs significantly, compared to the more typical BT that is based on secondary data.

While including contextual covariates does improve the temporal transfer for the test–retest respondents, their inclusion worsens the transfer across non-identical respondents. In line with the recommendations in for example Bateman et al. (2011), we also conducted a function transfer based only on theory-derived covariates. This resulted in a larger TE for the test–retest sample indicating that the statistically best-fit function performs better in terms of the size of TE. This is similar to findings in Whitehead and Hoban (1999), Berrens et al. (2000) and Zandersen et al. (2007), but at odds with the result found in Brouwer and Bateman (2005), Brouwer (2006) and Bateman et al. (2011).

In the test–retest part of our survey we see some changes in individuals' stated WTP from 2005 to 2010. While the typical standard assumption is that preferences are stable, there could be a range of reasons why an individual would change stated WTP over time. One explanation, which would have devastating consequences for the CVM method as such, could be that respondents do not take the CVM questions seriously and just answer more or less randomly. However, when modelling changes in WTP, we find several structurally significant explanatory variables that conform to theoretical predictions as the test–retest sample has actually changed slightly over time. Specifically, we find that latent factors such as perceived increases in risk of flooding and increasing global warming as well as increases in direct demand for better protection, all lead to increased WTP statements. Moreover, observable factors in terms of respondents who have increased their income or live in areas where flood risk is particularly high, also led to significant increases in WTP from 2005 to 2010.

Even though our results clearly suggest that benefit estimates can be reliably transferred over a 5-year period when no major preference-altering events have happened, we abstain from generalising based on this single empirical case. Considering the constantly increasing use of BT in policy and project evaluation and decision-making as well as the fact that BT will always involve a transfer over time, and in many cases over a time span of several years, there is an apparent need to further assess the temporal validity and reliability of BT. Specifically, further test–retest studies considering increasing time intervals would seem necessary to assess the expected lifetime of primary study benefit estimates and thus their reliability in practical BT.

Acknowledgments We gratefully acknowledge the Danish Coastal Authority for funding the distribution and collection of the surveys. Furthermore, we appreciate the helpful comments and suggestions from the two anonymous reviewers.

References

Allen BP, Loomis JB (2008) The decision to use benefit transfer or conduct original valuation research for benefit-cost and policy analysis. Contemp Econ Policy 26(1):1–12

Barton DN (1999) The quick, the cheap and the dirty: benefit transfer approaches to the non-market valuation of coastal water quality in Costa Rica. PhD Thesis no. 1999:03, Department of Economics and Social Sciences, Agricultural University of Norway

Barton DN (2002) The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica. Ecol Econ 42(1):147–164

Baskaran R, Cullen R, Colombo S (2010) Testing different types of benefit transfer in valuation of ecosystem services: New Zealand winegrowing case studies. Ecol Econ 69(5):1010–1022

Bateman IJ, Jones AP, Nishikawa N, Brouwer R (2000) Benefits transfer in theory and practice: a review and some new studies. http://www.uea.ac.uk/~e089. CSERGE and School of Environmental Sciences, University of East Anglia

Bateman IJ, Carson RT, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Özdemiroglu E, Pearce DW, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar, Cheltenham

Bateman IJ, Brouwer R, Ferrini S, Schaafsma M, Barton DN, Dubgaard A, Hasler B, Hime S, Liekens I, Navrud S (2011) Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe. Environ Resour Econ 50(3):365–387

Bergland O, Magnussen K, Navrud S (1995) Benefit transfer: testing for accuracy and reliability. Discussion Paper D-03/1995, Department of Economics and Social Sciences. Agricultural University of Norway

Berrens RP, Bohara AK, Silva CL, Brookshire D, Mckee M (2000) Contingent values for New Mexico in stream flows: with tests of scope, group-size reminder and temporal reliability. J Environ Manage 58(1):73–90

Bhattacharjee S, Kling CL, Herriges J (2009) A Kuhn–Tucker estimation of recreation demand—a study of temporal stability. In: Agricultural and applied economics association 2009 AAEA & ACCI joint annual meeting, July 2009. Milwaukee, Wisconsin

Bliem M, Getzner M (2012) Willingness-to-pay for river restoration: differences across time and scenarios. Environ Econ Policy Stud 14(3):1–20

Bliem M, Getzner M, Rodiga-Laßnig P (2012) Temporal stability of individual preferences for river restoration in Austria using a choice experiment. J Environ Manage 103:65–73

Brouwer R (2000) Environmental value transfer: state of the art and future prospects. Ecol Econ 32(1):137–152

Brouwer R (2006) Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event. Ecol Econ 60(2):399–406

Brouwer R, Spaninks F (1999) The validity of transferring environmental benefits: further empirical testing. Environ Resour Econ 14:95–117

Brouwer R, Bateman IJ (2005) Benefits transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study. J Health Econ 24(3):591–611

Cameron JI (1997) Applying socio-ecological economics: a case study of contingent valuation and integrated catchment management. Ecol Econ 23(2):155–165

Carlsson F, Kataria M, Krupnick A, Lampi E, Löfgren Å, Qin P, Chung S, Sterner T (2012) Paying for mitigation: a multiple country study. Land Econ 88(2):326–340

Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Presser S, Rudd PA, Smith VK, Conaway M, Martin K (1997) Temporal reliability of estimates from contingent valuation. Land Econ 73(2):151–163

Chattopadhyay S (2003) A repeated sampling technique in assessing the validity of benefit transfer in valuing non-market goods. Land Econ 79(4):576–596

Colombo S, Hanley N (2008) How can we reduce the errors from benefits transfer? An investigation using the choice experiment method. Land Econ 84(1):128–147

Downing M, Ozuna T Jr (1996) Testing the reliability of the benefit function transfer approach. J Environ Econ Manage 30(3):316–322

Dubgaard A, Olsen SB, Jespersen MML, Bonnichsen O, Klagenberg PA, Nielsen CCM (2011) Opfattelse af Risiko for Oversvømmelse 2010 [Perception of the risk of flooding 2010—in Danish only]. Joint report from Institute of Food and Resource Economics, University of Copenhagen, and the Danish Coastal Authority, Lemvig, Denmark

Dubourg WR, Jones-Lee MW, Loomes G (1994) Imprecise preferences and the WTP-WTA disparity. J Risk Uncertain 9(2):115–133

Eshet T, Baron MG, Shechter M (2007) Exploring benefit transfer: disamenities of waste transfer stations. Environ Resour Econ 37(3):521–547

Håkansson C (2008) A new valuation question: analysis of and insights from interval open-ended data in contingent valuation. Environ Resour Econ 39(2):175–188

Johnston RJ, Moeltner K (2013) Meta-modelling and benefit transfer: the empirical relevance of source-consistency in welfare measures. Environ Resour Econ. doi:10.1007/s10640-013-9730-3

Johnston RJ, Rosenberger RS (2009) Methods, trends and controversies in contemporary benefit transfer. J Econ Surv 24(3):479–510

Källstrøm MN, Hasler B, Olsen SB, Brodersen SL, Ørby PV, Levin G (2010) Testing different approaches to benefit transfers between two sites in the same country, valuing the improvement of water quality. Paper presented at the 11th biennial ISEE conference, 22–25 August 2010, Oldenburg and Bremen, Germany

Kaul S, Boyle KJ, Kuminoff NV, Parmeter CF, Pope JC (2013) What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent Validity. J Environ Econ Manage 66(1):90–104

Kealy MJ, Dovidio JF, Rockel ML (1988) Accuracy in valuation is a matter of degree. Land Econ 64(2): 158–171

Kirchhoff S, Colby BG, Lafrance JT (1997) Evaluating the performance of benefit transfer: an empirical inquiry. J Environ Econ Manage 33(1):75–93

Kristofersson D, Navrud S (2005) Validity tests of benefit transfer—are we performing the wrong tests? Environ Resour Econ 30(3):279–286

León CJ, Araña JE (2012) The dynamics of preference elicitation after an environmental disaster: stability and emotional load. Land Econ 88(2):362–381

Liebe U, Meyerhoff J, Hartje V (2012) Test–retest reliability of choice experiments in environmental valuation. Environ Resour Econ 53(3):1–19

Loomis JB (1989) Test–retest reliability of the contingent valuation method: a comparison of general population and visitor responses. Am J Agric Econ 71(1):76–84

Loomis JB (1992) The evolution of a more rigorous approach to benefit transfer: benefit function transfer. Water Resour Res 28(3):701–705

Mann HB, Whitney DR (1947) On a test whether one of two random variables is stochastically larger than the other. Ann Math Stat 18(1):50–60

McConnell KE, Strand IE, Valdés S (1998) Testing temporal reliability and carry-over effect: the role of correlated responses in test–retest reliability studies. Environ Resour Econ 12(3):357–374

Moeltner K, Rosenberger RS, (2012) Cross-context benefit Ttransfer: a Bayesian search for information pools. Paper presented at the 2012 EAERE annual meetings, 27–30 June 27–30. Czech Republic, Prague

Muthke T, Holm-Mueller K (2004) National and international benefit transfer testing with a rigorous test procedure. Environ Resour Econ 29(3):323–336

Navrud S (2007) Practical tools for value transfer in Denmark: guidelines and an example. Danish Environmental Protection Agency

Navrud S, Ready R (2007) Lessons learned for environmental transfer. In: Navrud S, Ready R (eds) Environmental value transfer: issues and methods. Springer, Netherlands

Onwujekwe O, Fox-Rushby J, Hanson K (2005) Inter-rater and test–retest reliability of three contingent valuation question formats in south-east Nigeria. Health Econ 14(5):529–536

Parsons GR, Stefanova S (2009) Temporal stability of recreation choices. In: 2009 Annual meeting, July 26–28, 2009, Milwaukee, Wisconsin (no. 49296), Agricultural and Applied Economics Association

Pearce D, Atkinson G, Mourato S (2006) Cost-benefit analysis and the environment: recent developments. OECD Publishing, France

Ready R, Navrud S, Dubourg R (2001) How do respondents with uncertain willingness to pay answer contingent valuation questions? Land Econ 77(3):315–326

Rozan A (2004) Benefit transfer: a comparison of WTP for air quality between France and Germany. Environ Resour Econ 29(3):295–306

Skjoldborg US, Lauridsen J, Junker P (2009) Reliability of the discrete choice experiment at the input and output level in patients with rheumatoid arthritis. Value Health 12(1):153–158

Smith VK, Pattanayak SK, Van Houtven GL (2006) Structural benefit transfer: an example using VSL estimates. Ecol Econ 60(2):361–371

Statistics Denmark (2012) http://www.dst.dk/en. Cited 18 Dec 2012

Teisl MF, Boyle KJ, Mccollum DW, Reiling SD (1995) Test–retest reliability of contingent valuation with independent sample pretest and posttest control groups. Am J Agric Econ 77(3):613–619

Tranberg J, Christoffersen LB, Dubgaard A, Olsen SB, Lassen C (2005) Opfattelse af Risiko for Oversvømmelse [Perception of the risk of flooding—in Danish Only]. University of Copenhagen, Report from the Institute of Food and Resource Economics

Verbeek M (2008) A guide to modern econometrics. Wiley, West Sussex

Whitehead JC, Hoban TJ (1999) Testing for temporal reliability in contingent valuation with time for changes in factors affecting demand. Land Econ 75(3):453–465

Wooldridge JM (2009) Introductory econometrics: a modern approach, 2nd edn. South-Western, Thomson Learning, USA

Wooldridge JM (2010) Econometric analysis for cross section and panel data. MIT Press, Cambridge

Zandersen M, Termansen M, Jensen FS (2007) Testing benefits transfer of forest recreation values over a twenty-year time horizon. Land Econ 83(3):412–440